Computationally efficient banding of large covariance matrices for ordered data and connections to banding the inverse Cholesky factor

Y. Wang
wang.yanpin@scrippshealth.org
Scripps Health
La Jolla, CA 92037

M. J. Daniels
mjdaniels@austin.utexas.edu
University of Texas at Austin
Austin, TX 78712

Summary

In this article, we propose a computationally efficient approach to estimate (large) p-dimensional covariance matrices of ordered (or longitudinal) data based on an independent sample of size n. To do this, we construct the estimator based on a k-band partial autocorrelation matrix with the number of bands chosen using an exact multiple hypothesis testing procedure. This approach is considerably faster than many existing methods and only requires inversion of $(k+1)$-dimensional covariance matrices. The resulting estimator is positive definite as long as $k < n$ (where p can be larger than n). We make connections between this approach and banding the Cholesky factor of the modified Cholesky decomposition of the inverse covariance matrix (Wu and Pourahmadi, 2003) and show that the maximum likelihood estimator of the k-band partial autocorrelation matrix is the same as the k-band inverse Cholesky factor. We evaluate our estimator via extensive simulations and illustrate the approach using high-dimensional sonar data.

Key words:

High dimensional covariance matrices; Hypothesis testing; Partial autocorrelations.

1 Introduction

Estimating a covariance matrix is essential in multivariate data analysis. Although the sample covariance matrix is an unbiased estimator of the covariance matrix of a Gaussian random vector, it has poor properties if the dimension (p) is large. In addition, covariance matrices are often sparse for large p. This area has seen an upsurge in practical and theoretical approaches due to a plethora of high dimensional data. Regularizing large sample covariance matrices has been proposed using ridge regression (Warton, 2008; Witten and Tibshirani, 2009) and discriminant analysis (Friedman, 1989). Furrer and Bengtsson (2007) consider 'tapering' the sample covariance matrix by gradually shrinking the off-diagonal elements toward zeros. Johnstone and Lu (2007) consider a regularization of principal components using a sparse basis and thresholding. Fan, Fan and Lv (2008) impose sparsity on the covariance matrix via a factor model.

Covariance matrices have special features under ordered (longitudinal) data, such as X_i and X_j being closer to independence or conditional independence as $|i - j|$ increases. Wu and Pourahmadi (2003) exploit this feature by banding the Cholesky factor of the inverse covariance matrix and use an information based approach (the AIC) to determine the number of bands. Bickel and Levina (2008) achieve sparsity by directly banding the sample covariance matrix while Rothman et. al. (2010) band the Cholesky factor of the covariance matrix. The former uses 'tapering' to maintain positive-definiteness (Furrer and Bengtsson, 2007). In addition, the latter two methods use cross-validation to find the number of bands, which can be computationally intensive. A nice review of high-dimensional covariance estimation can be found in Pourahmadi (2011).

In this article, we present a method for ordered/longitudinal data based on banding a different matrix, the partial autocorrelation matrix, which has favorable properties including the estimator guaranteed to be positive-definite (w/o any adjustment), even for $n < p$, and being based on exact, small sample results (not asymptotics). Computationally the estimator finds the number

of bands sequentially, so if the estimate has k bands, only $(k+1)$-dimensional matrices need to be inverted (not p-dimensional, where p is often large). There are also interesting connections to the banding approach in Wu and Pourahmadi (2003), which we point out here.

Our paper is arranged as follows. In section 2, we briefly review the partial autocorrelation matrix. In section 3, we introduce three new theoretical results which will allow us to develop an algorithm to easily estimate the non-zero elements in a banded partial autocorrelation matrix and develop a sequential multivariate hypothesis testing set-up (to estimate the number of bands), based on exact small sample results. Section 4 makes connections of our estimator to the banding estimator proposed in Wu and Pourahmadi (2003). Sections 5 and 6 investigate the operating characteristics of our procedure via risk simulations and apply it to two real data examples.

2 Review of Partial Autocorrelations

We first review reparameterizing the correlation matrix $R = (\rho_{j,k})$ using the elements in the partial autocorrelation matrix $\Pi = (\pi_{j,k})$. In the partial autocorrelation matrix, $\pi_{j,j} = 1$ and for $1 \leq j < k \leq p$, $\pi_{j,k}$ is the partial autocorrelation between Y_j and Y_k adjusted for the intervening variables, $\tilde{Y}_{jk} = (Y_{j+1}, \ldots, Y_{k-1})$. In particular, for a mean-zero Y, let $\tilde{Y}_{jk} = (Y_{j+1}, \ldots, Y_{k-1})^T$ be the vector of the intervening responses, and $\hat{Y}_j$ ($\hat{Y}_k$) be the linear least squares predictor of Y_j (Y_k) given $\tilde{Y}_{jk}$. The partial autocorrelation $\pi_{j,k}$ between Y_j and Y_k is $\pi_{j,k} = \text{corr}\{Y_j - \hat{Y}_j,\, Y_k - \hat{Y}_k\}$. The partial autocorrelations can be also be written directly as a function of the marginal correlations. We refer to the reader to Daniels and Pourhamadi (2009) for these recursive expressions.

In longitudinal data, we expect the matrix of partial autocorrelations to be sparse as the elements correspond to conditional independencies (for multivariate normal responses) and are an intuitive parameterization in such settings. An AR(1) correlation matrix corresponds to a partial autocorrelation matrix with one non-zero band with equal elements; however, the corresponding marginal correlation matrix is full. More generally the partial autocorrelation will

have k non-zero bands under a k-th order ante-dependence model (Gabriel, 1962; Zimmerman and Nunez-Anton, 2010). This parameterization has further advantages in that one can simultaneously address the partial autocorrelations and the marginal variances as opposed to related decompositions like the modified Cholesky decomposition which allows modeling of autoregressive coefficients and the prediction (or innovation) variances.

3 Banding the Partial Autocorrelation Matrix

For a $p \times p$ matrix $M = [m_{ij}]$ and any $k \in \{0, 1, 2, \ldots, p-1\}$, we define a k-band matrix of M (Bickel and Levina (2008)) as $B_k(M) = [m_{ij}\mathbf{I}(|i-j| \leq k)]$. Here we band the partial autocorrelation matrix, $\Pi = (\pi_{j,k})$. The computational attractiveness of our approach will rely on estimating the partial autocorrelations one band at a time and then for each band, doing a simple hypothesis test of whether to add another band. For the proposed procedure, the largest matrices we will need to manipulate will be $(k+1)$-dimensional matrices where k is the number of bands. In Section 3.1, we state the results necessary for the validity of our banding estimation approach (with proofs in the supplementary materials).

3.1 Key New Theoretical Results

In the following, we assume the data, $\{Y_i : i = 1, ..., n\}$ are independent and identically distributed (iid) multivariate normal p-vectors with mean 0 and covariance matrix Σ. We state and prove three key results that provide the foundation for our approach. The proofs of all three results can be found in the Supplementary materials. The first result states that the inverse of a correlation matrix constructed from a banded partial autocorrelation matrix only requires inversion of low dimensional matrices.

Result 1: Inverting the correlation matrix constructed from a k-band partial autocorrelation matrix only requires inversion of $(k+1)$-dimensional matrices, and its precision matrix is also a

k-band matrix.

The second result states that we can compute the maximum likelihood estimate (mle) of a k-band partial autocorrelation matrix one band at a time (starting from the first band). This will be much more efficient than having to manipulate the entire p-dimensional matrix. We can do this since the mle of the partial autocorrelations in band j only depends on the mle of the partial autocorrelations in the bands $< j$, not those $> j$.

In particular, we introduce a sequential procedure to estimate the partial autocorrelations in each band and show that is equivalent to the mle $(\hat{\Pi})$ of full multivariate normal likelihood $L(\Pi|y_1, ..., y_n)$. Let $f(y_{ij}, ..., y_{ik}|\Pi_{jk})$ be the marginal pdf of the random sub-vector $Y_i^{jk} = (Y_{ij}, ..., Y_{ik})$ of a multivariate normal random vector Y_i, and let $L(\Pi_{jk}|y_1^{jk}, ..., y_n^{jk})$ be the likelihood function based on this subvector of Y. We define an objective function $G(\pi_{jk})$ as follows:

$$G(\pi_{j,k}) \equiv L(\Pi_{jk}|y_1^{jk}, ..., y_n^{jk}),$$

where Π_{jk} denotes the subset $\{\pi_{j+1,j}, \pi_{j+2,j+1}, ..., \pi_{k,k-1}, \pi_{j+2,j}, ..., \pi_{k,j}\}$, for $k = 2, \ldots, p$, and $j = k-1, \ldots, 1$.

Since $\pi_{j,k}$ for $j, k \in \{1, 2, ..., p\}$ and $k \neq j$, independently vary in $(-1, 1)$ (see Result S2 in the supplementary materials), we can estimate $\pi_{1,2}, \pi_{2,3}, ..., \pi_{p-1,p}, \pi_{1,3}, ..., \pi_{p-2,p}, ..., \pi_{1,p}$ sequentially by maximizing the product of the objective functions in (3.1) for $l = 1, \ldots, p-1$,

$$\tilde{\pi}_{j,j+l} = \max_{\pi_{j,j+l}} G^{\star}(\pi_{j,j+l} : j = 1, \ldots, p-l), \tag{1}$$

where $G^{\star}(\pi_{j,j+l} : j = 1, \ldots, p-l) = \prod_{j=1}^{p-l} G(\pi_{j,j+l})$. The form of the estimating equations can be found in the supplementary materials. The maximizer of this objective function for each $\pi_{j,j+l}$ is equivalent to the mle of Π based on the multivariate normal likelihood which we state formally in the following theorem.

Theorem 1: The mle $\hat{\pi}$ of the partial autocorrelation coefficients based on the multivariate normal likelihood function $L(\Pi|y_1, y_2, ..., y_n)$ is equal to $\widetilde{\pi}$ in (1).

Since the estimated lag k partial autocorrelation coefficients are invariant to the estimated partial autocorrelation coefficients for lag greater than k, we obtain the following corollary.

Corollary 1: Let $\hat{\Pi}$ be the mle of partial autocorrelation matrix based on the multivariate normal likelihood function $L(\Sigma|y_1, y_2, ..., y_n)$ and $\widetilde{\Pi}$ based on (1). Then the mle $(\hat{\Pi})$ of a k-band partial autocorrelation matrix is equivalent to the corresponding $\widetilde{\Pi}$ of a k-band matrix.

This result allows us to compute the mle of a k-band Π efficiently (actually by just solving quadratic equations; see the supplementary materials for details).

The third result is the final piece needed to do an exact sequential hypothesis testing procedure to estimate the number of bands as it provides the sampling distribution of the partial autocorrelations in bands $j \geq k+1$ of a $k-$band matrix.

Theorem 2: Suppose $Y_1, Y_2, ..., Y_n$ are iid $N(0, DRD)$ and $\Pi = (\pi_{j,t})_{p\times p}$ is the partial autocorrelation matrix of R. Then, for $n \geq p+1$, under the hypothesis of a k-band matrix, the mle's of $\pi_{j,j+l}$, denoted as $\hat{\pi}_{j,j+l}$, $l = 1, \ldots, p-1$ follow independent shifted Beta distributions on $(-1, 1)$ with parameters

$$\alpha = \beta = \begin{cases} \frac{n-l-2}{2} & \text{if } 1 \leq k \leq l \leq p-2 \\ \frac{n-p-1}{2} & \text{if } k \leq l = p-1 \end{cases}.$$

3.2 Procedure to estimate the partial autocorrelation matrix

We use the three results from the previous section to construct a procedure for banding the partial autocorrelation matrix. To estimate the number of bands (k) of a partial autocorrelation matrix $\Pi = (\pi_{j,k})$, the strategy will be sequentially testing the null hypothesis that each band is zero starting from the first band. Implicitly, if the jth band is zero, the subsequent bands, $j+1, \ldots, p-1$ are zero as well. In general, for $\pi_k = (\pi_{1,1+k}, ..., \pi_{p-k,p})$, we construct multiple

tests under the following hypotheses:

$$H_0 : \pi_l = 0, l \geq k \text{ vs } H_a : \pi_k \neq 0.$$

We choose the band as the first k for which H_0 can not be rejected. Note we just need to test the partial autocorrelations in the kth band under the assumption of a true band $k-1$ matrix. Since under the null hypothesis, the partial autocorrelations are independent (shifted) Beta distributions as given in Theorem 2, we adjust for multiple testing using a Bonferroni correction to account for the $(p-k)$ tests in band k. In addition, from Result 1 we only need to manipulate $(k+1)$ dimensional matrices for a k-band matrix.

For $n \leq p$ and lag l $(k > l)$, all $(p-l)$ sample partial autocorrelations are not independent, only each set of $n-l$ adjacent ones. Based on empirical checks, the correlations appear to very small and as a result, we still use the Bonferroni correction. We explore the operating characteristics of this procedure via simulations in Section 5.

4 Connections with Banding the inverse Cholesky factor

Wu and Pourahmadi (2003) proposed banding the (modified) Cholesky factor of the inverse covariance matrix (Pourahmadi, 1999). We briefly review the details next. As before, we assume the p-dimensional vector Y is distributed as $Y \sim N(0, \Sigma)$. As such, it is easy to show that

$$Y_j = \sum_{t=1}^{j-1} a_{jt} Y_t + \epsilon_j = Z_j^T a_j + \epsilon_j,$$

where $Z_j^T = (Y_1 ..., Y_{j-1})$, $a_j^T = (a_{j1}, \ldots, a_{j,j-1})$, and $\epsilon_j \sim N(0, \sigma_j^2)$. The coefficients and variances here are the parameters in the modified Cholesky decomposition given as $A\Sigma A^T = D$, where A is a unit lower triangular matrix having ones on its diagonal and $-a_{jt}$ as its (j,t)th element (for $t < j$), and D is a diagonal matrix with $\sigma_j^2 = var(\epsilon_j)$ as its diagonal entries

(Pourahmadi, 1999). The parameters, a_{jt} and σ_j^2, are often called generalized autoregressive parameters and innovation variances, respectively.

From linear model theory, it follows that for a k-band matrix, the mle of the components of the Cholesky factor, $\hat{a}_{jt}$ are independently distributed within bands (but not across bands) as

$$(\hat{a}_{jt}^{(k)} - a_{jt})/\sqrt{\hat{\sigma}_j^{2(k)} s_j^{tt}} \sim t_{n-\min(k,j-1)},$$

where s_j^{tt} is the jth diagonal entry of $(Z_j^{(k)} Z_j^{(k)T})^{-1}$, $Z_j^{(k)} = (Y_{j-k}, \ldots, Y_{j-1})$, and $\sigma_j^{2(k)}$ is the mle of the innovation variances for the k-band matrix.

Given this result, to estimate the number of bands (k), we can sequentially test the null hypothesis that each band of A is zero starting from the first band just as we did for the partial autocorrelation matrix. Since the components in each band of the Cholesky factor are independent t-distributions, we again adjust for multiple testing using a Bonferroni correction. Note that, unlike with the partial autocorrelations, when we reject the null hypothesis and move to the next band, we need to recompute all the estimates of the generalized autoregressive parameters from the previous bands.

Wu and Pourahmadi (2003) chose the number of bands (k) using the following form of the Akaike Information Criterion (AIC),

$$\text{AIC} = \begin{cases} n \sum_{t=1}^{p} \log \hat{\sigma}_t^{2(k)} + 2(p - k/2)(k+1), & k = 1, \ldots, \min(p-1, n-2), \\ n \sum_{t=1}^{p} \log\{\frac{\sum_{i=1}^{n}(y_{it} - \bar{y}_t)^2}{n}\} + 2p, & k = 0, \end{cases}$$

where $\bar{y}_t = (1/n) \sum_{i=1}^{n} y_{it}$ and $\hat{\sigma}_t^{2(k)} = \hat{\epsilon}_t^{(k)T} \hat{\epsilon}_t^{(k)}/n$ with $\hat{\epsilon}_t^{(k)} = (I_n - P_t^{(k)}) Y_t$ and $P_t^{(k)} = Z_t^{(k)T} (Z_t^{(k)} Z_t^{(k)T})^{-1} Z_t^{(k)}$. They set the maximum number of bands to be $\lfloor p^{1/3} \rfloor$.

We now state a result that provides insight in using the (AIC) criterion for banding the partial autocorrelation matrix or the inverse Cholesky factor.

Result 2: The mle of the k-band modified Cholesky decomposition is equivalent to the mle of the

k-band partial autocorrelation matrix when combined with the mle's of the marginal variances.

Given Result 2, we can use this same criterion for banding the partial autocorrelation matrix by forming the corresponding covariance matrix using the mle's of the marginal variances. We compare the sequential hypothesis testing to the AIC criterion in the simulations.

5 Simulations

To evaluate our banding method, we conducted simulations using several true matrices. Scenario 1 was an AR(1) correlation matrix with lag 1 correlation equal to 0.7 (1 band). Scenario 2 was a matrix formed by banding (4, 9, or 14 bands) the sample partial autocorrelation matrix of the Metal data (for $p = 60$); for more details on the data, see Section 6. Note the true matrices in scenario 2 were not stationary or smooth. For scenario 1, we considered: 1) fixed dimension matrices ($p = 60$) with varying sample sizes (n), 2) fixed sample size ($n = 100$) with varying dimension (p). We generated 100 replicated datasets for each scenario from a multivariate normal distribution with mean zero and covariance matrix given by the correlation matrices above.

We compared our estimator (using $\alpha = 0.05$) to the sample correlation matrix ($\hat{R}$), the sequential hypothesis testing using the modified Cholesky decomposition ($\alpha = .05$), and an information based approach (AIC in Section 4) based on the mle of the covariance matrix formed from the k-band partial autocorrelation matrix; note we did not use the recommendation of Wu and Pourahmadi that the maximum band should be $\lfloor p^{1/3} \rfloor$ as this resulted in very poor performance for several scenarios. We used the average Frobenius matrix norm of the difference between estimated and true correlation matrices over the replicated datasets to compare the estimators. The results here using Frobenius norm were similar to those using other matrix norms. The estimated number of bands (k) was obtained by averaging the estimated bands over the replicated datasets.

Tables 1 and 2 summarize the simulation results. In Table 1 (scenario 1), the risk results indicated that all the estimators had much lower risk than the sample correlation matrix (which

was not surprising). Also, the risks and the estimated number of bands were very similar among the two sequential testing approaches and the AIC approach with the AIC approach doing the best by a very small margin.

Table 2 contains the simulation results for scenario 2, which corresponded to a nonstationary correlation matrix with 4, 9 or 14 bands. Again, all the proposed estimators had much smaller risk than sample correlation matrix. Among the two sequential testing estimators, the results were essentially indistinguishable. And compared to the AIC approach, the AIC approach generally resulted in a slightly smaller risk and was able to detect more bands for the small sample size cases. The lack of independence of all $(p-k)$ tests in each band appear to have minimal influence on the results as both sequential testing approaches gave very similar results. Other simulation scenarios with similar conclusions can be found in the supplementary materials. In summary, all three approaches performed quite similarly.

6 Application to Sonar Data

We illustrate our approach on two data sets, the Metal and Rock data of the sonar data, which is available at $http : www.ics.uci.edu/\tilde{}melearn/MLRepository.html$. This data set contains 111 (97) signals from a Metal cylinder (Rock), where each signal has 60 frequency energy measurements ranging from 0.0 to 1.0. These signals were measured at different angles for the same objects. Previous analysis of the data assumed the signals were iid normal random vectors. Images of absolute sample correlation matrices of the data in Figure 1 (hot to cool corresponds to 1.0 to 0.0) indicate a general pattern of decaying correlations with increasing lag, which motivates the banded estimator here. The estimated number of bands using both sequential testing procedures was 3 for rock data and 6 for metal data, respectively. For the information based banding approach, the number of bands was 4 and 11, respectively. The images of estimated correlation matrices based on these estimators are shown in Figure 1. Subfigures (a) – (c) for

the metal data show that most of the nonzero marginal correlations up to lag 31 are captured quite well and Subfigures (d) – (f) for the Rock data show similar results up to lag 18. All the estimators are computationally very fast, taking less than one second with Matlab R2010a.

7 Discussion

We have proposed k-band estimators for a correlation matrix that are positive definite even when $n \leq p$ and whose computation only requires inversion of at most $(k+1)$-dimensional matrices. The algorithm for the estimator relies on exact distributional results under the null hypothesis for $n > p$. The estimator can be computed very quickly.

We have made connections between banding the partial autocorrelation matrix and the modified Cholesky factor of the inverse covariance matrix (Wu and Pourahmadi, 2003). Sequential hypothesis testing using either factorization results in essentially the same results. In addition, implementing the AIC approach of Wu and Pourahmadi appears to do marginally better (and gives an equivalent result for banding the partial autocorrelation matrix or Cholesky factor) once the upper bound of $\lfloor p^{1/3} \rfloor$ on the number of bands was removed. The key with these approaches is to start with a zero band matrix and move sequentially until stopping (based on not rejecting the relevant null hypothesis or the AIC increasing). The sequential hypothesis testing under the partial autocorrelations has some computational advantage as the estimates from previous bands do not need to be recomputed as the number of bands increases.

The modified Cholesky factorization and the variance/correlation decomposition (implicit in using the partial autocorrelation matrix) correspond to different sets of dependence parameters (generalized autoregressive parameters vs. partial autocorrelations) and variance parameters (innovation vs. marginal variances). The latter might be preferred for smoothing bands one at a time as the interpretation of the partial autocorrelations in the jth band does not depend on autocorrelations in bands larger than j unlike the generalized autoregressive parameters. In addition,

it is often preferred to smooth marginal variances (for interpretation) than innovation variances.

Acknowledgments

Work was supported by NIH grant CA85295.

Supplementary Materials

Supplementary materials include proofs of results and theorems in the main text and additional theoretical results and simulations.

References

[1] Peter J. Bickel and Elizaveta Levina. Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics*, 36(1):199–227, 2008.

[2] Michael J. Daniels and Mohsen Pourahmadi. Modeling Covariance Matrices via Partial Autocorrelations. *Journal of Multivariate Analysis*, 100:2352–2363, 2009.

[3] Jianqing Fan, Yingying Fan, and Jinchi Lv. High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics*, 147(1):186–197, 2008.

[4] Jerome H. Friedman. Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):pp. 165–175, 1989.

[5] Reinhard Furrer and Thomas Bengtsson. Estimation of High-dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants. *Journal of Multivariate Analysis*, 98:227–255, 2007.

[6] K. R. Gabriel. Ante-dependence Analysis of an Ordered Set of Variables. *The Annals of Mathematical Statistics*, 33(1):pp. 201–212, 1962.

[7] Iain M Johnstone and Arthur Yu Lu. Sparse Principal Components Analysis. *arXiv preprint arXiv:0901.4392*, 2009.

[8] Mohsen Pourahmadi. Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterisation. *Biometrika*, 86(3):677–690, 1999.

[9] Mohsen Pourahmadi. Covariance Estimation: The GLM and Regularization Perspectives. *Statistical Science*, 26(3):369–387, 2011.

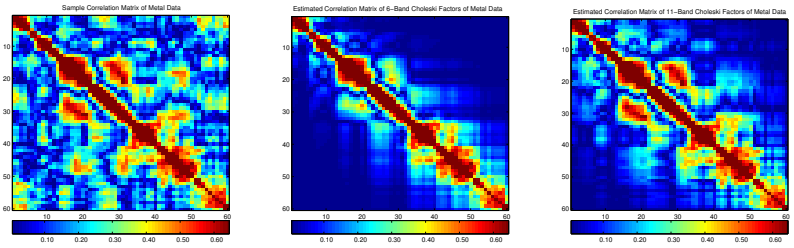
[10] Adam J. Rothman, Elizaveta Levina, and Ji Zhu. A new Approach to Cholesky-based Covariance Regularization in High Dimensions. *Biometrika*, 97(3):539–550, 2010.

[11] David I Warton. Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices. *Journal of the American Statistical Association*, 103(481):340–349, 2008.

[12] Daniela M. Witten and Robert Tibshirani. Covariance-Regularized Regression and Classification for High Dimensional Problems. *Journal Of The Royal Statistical Society Series B*, 71(3):615–636, 2009.

[13] Wei B. Wu and Mohsen Pourahmadi. Nonparameteric Estimation of Large Covariance Matrices of Longitudinal Data. *Biometrika*, 90(4):831–844, 2003.

[14] Dale L. Zimmerman and Vicente Nunez-Anton. *Antedependence Models for Longitudinal Data*. Number 11. Boca Raton, Florida: CRC Press, 2010.



(a) **(b)** **(c)**

Figure 1: Panels a and d are the images of sample correlation matrices of Metal and Rock data, respectively. Panels b and c (e and f) are images of estimated correlation matrices of metal (rock) data for 3 (6) bands (estimated by sequential testing) and 4 (11) bands (estimated by AIC).

p=60	$\hat{R}$		AIC		Chol-Testing		PAC-Testing	
Size (n)	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	10.76 (0.20)	1.00	2.29 (0.08)	1.09	2.36 (0.15)	1.09	2.36 (0.15)
60	7.49 (0.11)	1.00	1.61 (0.06)	1.04	1.63 (0.07)	1.04	1.63 (0.07)
100	5.87 (0.08)	1.00	1.24 (0.03)	1.06	1.26 (0.04)	1.06	1.26 (0.04)
n=100	$\hat{R}$		AIC		Chol-Testing		PAC-Testing	
Dim (p)	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
30	2.82(0.07)	1.00	0.82 (0.03)	1.01	0.82 (0.03)	1.01	0.82 (0.03)
100	9.83(0.08)	1.00	1.62 (0.03)	1.07	1.66 (0.06)	1.06	1.65 (0.05)
200	19.87(0.06)	1.00	2.28 (0.03)	1.06	2.33 (0.08)	1.06	2.33 (0.08)

Table 1: Results for scenario 1

Note: Frobenius matrix norms (Risk) and Monte Carlo variance (Var) over 100 replicates based on the banded partial autocorrelation estimator by multiple hypotheses testing (PAC-Testing), banded Cholesky factors of inverse covariance matrix by multiple hypotheses testing (Chol-Testing) or (AIC) for scenario 1. $\rho = 0.7$. EB is the estimated number of bands.

p=60		$\hat{R}$		AIC		Chol-Testing		PAC-Testing
Size(n)	Band	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)	EB	Risk (Var)
	4	10.74(0.39)	3.59	4.51(0.34)	3.35	4.53(0.38)	3.30	4.55(0.40)
30	9	10.30(0.64)	4.78	7.76(1.18)	3.83	8.08(1.06)	3.45	8.27(0.87)
	14	10.15(0.69)	5.04	9.53(0.98)	3.78	9.78(0.72)	3.49	9.96(0.59)
	4	7.40(0.21)	3.97	3.14(0.19)	3.96	3.18(0.19)	3.95	3.17(0.19)
60	9	7.33(0.33)	8.46	4.96(0.55)	7.29	5.33(1.04)	7.11	5.37(1.17)
	14	7.23(0.47)	10.69	6.24(1.53)	7.56	7.19(2.26)	7.32	7.27(2.37)
	4	5.76(0.10)	4.00	2.35(0.08)	4.04	2.36(0.08)	4.04	2.36(0.08)
100	9	5.65(0.18)	9.00	3.70(0.18)	8.71	3.87(0.41)	8.71	3.87(0.41)
	14	5.51(0.19)	13.62	4.35(0.19)	12.00	4.62(0.85)	11.98	4.61(0.86)
	4	2.56(0.03)	4.00	1.07(0.02)	4.05	1.07(0.02)	4.05	1.07(0.02)
500	9	2.52(0.05)	9.00	1.63(0.04)	9.01	1.63(0.04)	9.01	1.63(0.04)
	14	2.48(0.04)	14.00	1.94(0.04)	14.08	1.94(0.04)	14.07	1.94(0.04)

Table 2: Results for scenario 2

Note: Frobenius matrix norms (Risk) and Monte Carlo variance (Var) over 100 replicates based on the banded partial autocorrelation estimator by multiple hypotheses testing (PAC-Testing), banded Cholesky factors of inverse covariance matrix by multiple hypotheses testing (Chol-Testing) or (AIC) for scenario 2. EB is the estimated number of bands.